Global Change Biology

Global Change Biology (2016), doi: 10.1111/gcb.13376

Do maize models capture the impacts of heat and drought stresses on yield? Using algorithm ensembles to identify successful approaches

ZHENONG JIN[1], QIANLAI ZHUANG[1,2,3], ZELI TAN[1], JEFFREY S. DUKES[3,4,5], BANGYOU ZHENG[6] and JERRY M. MELILLO[7]

[1]*Department of Earth, Atmospheric and Planetary Sciences, Purdue University, West Lafayette, IN 47907, USA,* [2]*Department of Agronomy, Purdue University, West Lafayette, IN 47907, USA,* [3]*Purdue Climate Change Research Center, Purdue University, West Lafayette, IN 47907, USA,* [4]*Department of Forestry and Natural Resources, Purdue University, West Lafayette, IN 47907, USA,* [5]*Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA,* [6]*CSRIO Agriculture, Queensland Bioscience Precinct, 306 Carmody Road, St. Lucia, Qld 4067, Australia,* [7]*The Marine Biological Laboratory, The Ecosystems Center, Woods Hole, MA 02536, USA*

Abstract

Stresses from heat and drought are expected to increasingly suppress crop yields, but the degree to which current models can represent these effects is uncertain. Here we evaluate the algorithms that determine impacts of heat and drought stress on maize in 16 major maize models by incorporating these algorithms into a standard model, the Agricultural Production Systems sIMulator (APSIM), and running an ensemble of simulations. Although both daily mean temperature and daylight temperature are common choice of forcing heat stress algorithms, current parameterizations in most models favor the use of daylight temperature even though the algorithm was designed for daily mean temperature. Different drought algorithms (i.e., a function of soil water content, of soil water supply to demand ratio, and of actual to potential transpiration ratio) simulated considerably different patterns of water shortage over the growing season, but nonetheless predicted similar decreases in annual yield. Using the selected combination of algorithms, our simulations show that maize yield reduction was more sensitive to drought stress than to heat stress for the US Midwest since the 1980s, and this pattern will continue under future scenarios; the influence of excessive heat will become increasingly prominent by the late 21st century. Our review of algorithms in 16 crop models suggests that the impacts of heat and drought stress on plant yield can be best described by crop models that: (i) incorporate event-based descriptions of heat and drought stress, (ii) consider the effects of nighttime warming, and (iii) coordinate the interactions among multiple stresses. Our study identifies the proficiency with which different model formulations capture the impacts of heat and drought stress on maize biomass and yield production. The framework presented here can be applied to other modeled processes and used to improve yield predictions of other crops with a wide variety of crop models.

Keywords: crop model comparison, maize, yield, heat stress, drought stress, Agricultural Production Systems sIMulator

Received 21 December 2015; revised version received 25 March 2016 and accepted 4 May 2016

Introduction

The long-lasting and pervasive 2012 heat wave and drought in the United States damaged a substantial proportion of crop commodities, especially those in the Midwest (Mallya *et al*., 2013). Such an extreme climatic event (ECEs), however, is only a microcosm of the past decades full of fierce weather extremes (Coumou & Rahmstorf, 2012). These ECEs are projected to continue in the future, with increasing magnitude, duration, and frequency (IPCC, 2012). The rising incidence of weather extremes will exacerbate negative impacts on the crop productivity; indeed, critical thresholds are already being exceeded (Hatfield *et al*., 2014). As many other crops, contemporary maize production is threatened by the changing climate that reduces maize farming efficiency (Bassu *et al*., 2014). Concerns have thus been raised about maintaining a stable increase rate of the US maize yield, which is vital to the global food security (Bruinsma, 2009; Ort & Long, 2014). Extreme heat and drought are the two dominant constraints to the rainfed maize cultivating system in the United States (Schlenker & Roberts, 2009; Lobell *et al*., 2013; Hatfield *et al*., 2014).

Process-based crop models that incorporate maize modules are powerful tools for evaluating the potential impacts of climate change on maize yield (Bassu *et al*., 2014). Combined with hyper-local growth monitoring

Correspondence: Qianlai Zhuang, tel. +1 765 494 9610, fax +1 765 496 1210, e-mail: qzhuang@purdue.edu

and assimilation of high-resolution and real-time weather data, crop models can increasingly help stakeholders predict maize production and make decisions. However, these models remain poorly suited to manage and alleviate the risks from ECEs such as heat and drought. Most current generations of ecosystem models, including crop models, were originally optimized to simulate average conditions based on long-term climatology (Reichstein *et al.*, 2013), and their algorithms that simulate specific stresses are not well parameterized, either due to a lack of natural and experimental records of maize yield responses to high temperature and severe drought with which to train models, or due to a slow pace of updating model parameterizations. While broad agreement exists in terms of the effects of heat and drought on maize growth and development, researchers have abstracted this knowledge into markedly different equations and interactions (Saseendran *et al.*, 2008; Bassu *et al.*, 2014). Differences among algorithms are more prominent for heat than for drought, likely because fewer high-quality datasets have been available to describe heat stress effects on maize biomass production, grain set, grain fill, and yield. There is a clear need to systematically assess the environmental responses of biological processes in crop models, especially those processes that directly determine the simulated crop productivity.

As a critical first step toward model improvement, crop model comparison studies have become popular, especially for climate change scenarios (Rosenzweig *et al.*, 2013). In a review of 5 major crop models, Saseendran *et al.* (2008) found that these models all use the ratio of actual to potential transpiration or evapotranspiration to indicate water stress, but none of them can accurately represent the coupled processes of carbon assimilation, transpiration, energy balance, and stomatal behavior. Eitzinger *et al.* (2013) compared responses to heat and drought stress of seven widely used crop models and pointed out that even though a general consensus can be reached on the yield trend in response to increased temperatures, these models were not able to capture the direct heat stress impacts that account for substantial yield variations. More recently, the Agricultural Model Intercomparison and Improvement Project (AgMIP) has significantly advanced this field under protocols of coordinated evaluation, intercomparison, and improvement of crop models (Rosenzweig *et al.*, 2013). Asseng *et al.* (2013) observed that variations among crop models account for a greater proportion of the uncertainty in simulating global wheat yields under climate change than variations among future climate scenarios. By evaluating the performance of 23 maize models under four production conditions, Bassu *et al.* (2014) found that an ensemble of models was more reliable than one single model in capturing the mean yield even with very limited data for model calibration.

These comprehensive assessments advance the operational use of available crop models and shed light on the capability and uncertainty in the tools, but their findings often give only vague guidance to support individual model improvement (Donatelli *et al.*, 2014). This trade-off is inevitable in studies that compare the output from full models. As crop models differ substantially in the way they simulate crop physiology, soil physical characteristics and nutrient states, not to mention the differences in input variables and parameter settings, model developers often find it hard to tell which part of their models need to be improved when simply looking at the final results (e.g., yields). Some might argue that modelers can trace sources of uncertainty by examining intermediate variables, for instance by comparing leaf area index (LAI) with observations. Unfortunately, though, any of these intermediate variables themselves are results of complicated interactions among processes within a model. For a specific crop process (e.g., photosynthesis, phenology, and yield formation), there usually exist a number of ways to construct the mathematical algorithms (Bassu *et al.*, 2014; Martre *et al.*, 2015). Thus, to quantitatively understand the uncertainty related to that particular process, comparison should be performed in a way similar to a controlled experiment, such that any other processes are isolated.

The idea of focusing on different algorithms or different implementations of the same algorithm for a particular process (defined as 'algorithm ensemble' hereafter) when comparing crop models has been tested a few times and proved to be promising for elucidating the target issue (Saseendran *et al.*, 2008; Eitzinger *et al.*, 2013; Alderman *et al.*, 2014; Donatelli *et al.*, 2014; Kumudini *et al.*, 2014). It is favorable also because research advances that can be easily assimilated into models are mostly those at the process level (Donatelli *et al.*, 2014). However, very few studies have performed comparisons in a fully controlled style such that a process ensemble was quantitatively evaluated within a single platform (but see Donatelli *et al.* (2014) for a pioneering case study on soil temperature simulation). Insufficient modularization, poor documentation of most crop models and intellectual property boundaries are believed to be the three vital obstacles that hinder reimplementation and reuse of alternative algorithms for a specific process (Holzworth *et al.*, 2015).

In this study, we implement the 'algorithm ensemble' framework to evaluate the performance of difference algorithms in capturing the impact of heat and drought stress on maize biomass production and yield formation. We first review existing algorithms at the equation level from 16 major crop models that simulate the direct heat and drought stress on maize photosynthesis and yield formation, and document them for crop modelers (Data S1 in the Supporting Information). Next, we describe how representative algorithms were extracted from their parent crop models and incorporated into a standard model so that variations among algorithms could be quantified in a controlled manner. We select the Agricultural Production Systems sIMulator (APSIM) as the standard model, because its generic and modularized design allows algorithms to be replaced without changing the model structure. Finally, the revised APSIM with algorithm ensemble is used to simulate maize production at typical farms in the US Corn Belt, and evaluated using the county-level yield statistics from the USDA National Agricultural Statistics Service (NASS). Our goal is to understand why a particular algorithm (if any) performed better than others in capturing the signal of heat and drought, and to offer clear and useful information regarding crop model improvement. We exclude the evaluation of algorithms of indirect heat and drought stresses via leaf elongation/senescence, which are often programed to be more susceptible to adverse growth conditions (e.g., water stress effect in CERES-Maize), because the complex interactions between phenology and photosynthesis will make the results too complicated to interpret. We focus on maize because it is the most important cereal commodity in the United States, but the framework presented in our study can be extended to other crops and any process in a crop model.

Materials and methods

In this section, we first describe a method to quickly screen the behavior of heat stress algorithms. Next, we describe simulations that use an algorithm ensemble for the historical period of 1980–2013 and future scenarios of 2006–2099. A brief introduction to the development and application of APSIM-Maize model and its important engineering features is provided in Data S2. Screening of heat stress algorithms was conducted at the AmeriFlux Mead Rainfed station, Saunders, NE (41.18°, −96.44°), where hourly meteorological and fluxes variables were archived. Screening of drought stress algorithms was performed at Agricultural Engineering and Agronomy Research Farms of Iowa State University, Boone, IA (42.02°, −93.78°). The ensemble simulation was conducted at the Iowa farm as well as at two other sites: (i) the AmeriFlux Bondville station, Champaign, IL (40.01°, −88.29°); (ii) Purdue Agronomy Center for Research and Education, West Lafayette, IN (40.47°, −86.99°). For brevity, we mainly focus on the Indiana farm for the ensemble simulation, while present similar results from the other two farms in the Supporting Information.

Screening of stress functions

To understand the behavior of representative heat and drought response functions, we pulled out these algorithms from their parent models and reprogrammed each in R language. Such a 'lightweight' method allowed fast screening of these algorithms, while avoiding the 'heavy' task of running crop models, which usually requires extensive preparation.

For heat stress, we selected the temperature response curve of photosynthesis/carbon assimilation from AgroIBIS (Quadratic; Kucharik & Brye, 2003), APSIM (piecewise linear; Keating *et al*., 2003), CERES (piecewise linear; Jones *et al*., 2003), DayCent (Generalized Poisson; Parton *et al*., 1998), EPIC (Sinusoidal; Sharpley & Williams, 1990), MAIZSIM (Exponential; Yang *et al*., 2009), SWAT (Exponential; Neitsch *et al*., 2011), and WOFOST (piecewise linear; Supit *et al*., 1994). These 8 representative selections cover all different shapes of temperature response curves for the 16 crop models reviewed in this study (Table 1) and detailed descriptions for each can be found in Data S3. These temperature response curves were compared to the observed ratio of gross primary production (GPP) to absorbed photosynthetically active radiation (APAR) at different temperatures from the AmeriFlux Mead Rainfed station (Data S4). Next, we calculated the mean annual heat stress factors by integrating daily values over either the growing season. Daily weather inputs, including maximum and minimum air temperature at a spatial resolution of 1 km × 1 km, were downloaded from the Daymet website (http://daymet.ornl.gov/). During our preliminary analysis, we observed that models such as DayCent, SWAT, and WOFOST that use daily mean temperature to force the heat stress algorithm predicted almost no heat stress on annual basis, while the CERES model that uses daylight temperature (approximated by $\frac{T_{\max}+T_{\mathrm{mean}}}{2}$ hereafter) was more sensitive to excessive heat. Therefore, we also tested the effect of using daylight temperature to simulate heat stress. The simulation results were compared to growing season extreme degree days (EDD, which is cumulative daily mean of hourly temperature above 30 °C; Lobell *et al*., 2013) and killing degree days (KDD, which is the cumulative daily mean temperature above 29 °C; Butler & Huybers, 2013), both of which are indicators of excessive heat for crops (details of our implementation are given in Data S5).

For drought stress, we evaluated the three dominant algorithms that cover more than 80% of the crop models we reviewed (Table 2): functions of average soil moisture content (SWC), water supply to demand ratio (Ws/Wd), and actual to potential transpiration ratio (AT/PT). It should be noted that although Ws is close to AT because soil water supply largely determines the actual transpiration in many models, the denominators of Wd and PT are quite different, such that the former is based on the concept of transpiration efficiency (Hammer *et al*., 2010) and the latter directly reflects daily weather conditions (Allen *et al*., 1998). For simplicity, we used the APSIM SoilWat module (a tipping-bucket model) to simulate daily state variables and fluxes that were not directly observed. We calculated mean annual drought stress factors by averaging daily values over the growing season for each year. To reduce the uncertainty in hydrological modeling, we

Table 1 Summary of heat stress algorithm on maize photosynthesis, grain set/fillings, and harvest index. Detailed descriptions are given in Data S3

Model	Process	Model type	Input temperature	Key parameters	References
AgroIBIS	Stomatal resistance	Quadratic	Tleaf	Topt* = 25	Kucharik & Brye (2003)
APSIM	RUE	Multilinear	Tmean	Tbase† = 8, Topt1 = 15, Topt2 = 30, Tlim‡ = 44	Keating *et al.* (2003)
	Grain number	Linear	Tmax	Tlim = 38, Sensitivity = 0.05	Carberry *et al.* (1989)
	Grain filling	Linear	Tmean	Tcrt§ = c(6, 10, 16, 22, 30, 56.3)	
AquaCrop	Harvest index	Logistic	Tmean	Topt2 = 30, Tlim = 35	Raes *et al.* (2009)
CERES-4.0	RUE	Multilinear	Teff	Tbase = 6.2, Topt1 = 16.5, Topt2 = 33, Tlim = 44	Jones *et al.* (2003)
	Grain filling	Multilinear	Tmean	Tbase = 5.5, Topt1 = 16, Topt2 = 39, Tlim = 48.5	
CropSyst	Flowering	Multilinear	Thr	Tcrt = 31, Tlim = 44	Stockle *et al.* (2014)
DayCent	GPP	GPoisson	Tsoil	Topt = 30, Tlim = 45, Sleft = 1, Sright = 2.5	Parton *et al.* (1998)
EPIC	RUE	Sinusoidal	Tground	Tbase = 8, Topt = 25	Sharpley & Williams (1990)
GLAM	Flowering	Multilinear	Tam	To be calibrated	Challinor *et al.* (2005)
	Transpiration efficiency	Multilinear	Tmean	Tcrt = 35, Tlim = 47	Challinor *et al.* (2009)
HYBRID-maize	Assimilation rate	Multilinear	Tdaytime	Tbase = 8, Topt1 = 18, Topt2 = 30	Yang *et al.* (2013)
	Grain filling	Quadratic	T3 hour	Topt = 26	
CSM-IXIM	Assimilation rate	Complex			Lizaso *et al.* (2005)
MAIZSIM	Carbon supply	Exponential	Thr	Td = 48.6	Yang *et al.* (2009)
	Grain filling	Quadratic	Thr	Topt = 26	Grant (1989)
MONICA	Assimilation rate	Multilinear	Thr		Sage & Kubien (2007)
	Flowering	Multilinear	Tdaytime	Tcrt = 30, Tlim = 40	Moriondo *et al.* (2011)
PEGASUS	LUE	Quadratic	Tmean	Tbase = 0, Topt1 = 15, Topt2 = 40, Tlim = 65	Deryng *et al.* (2011)
	Flowering	Multilinear	Teff	Tcrt = 32, Tlim = 45	Deryng *et al.* (2014)
STICS	RUE	Quadratic	Tleaf	Tbase = 2.5, Topt1 = 10, Topt2 = 30, Tlim = 30	Brisson *et al.* (2009)
	Grain filling	Multilinear	Tleaf	Tbase = 5, Topt1 = 6, Topt2 = 26.5, Tlim = 27.5	
SWAT	RUE	Exponential	Tmean	Tbase = 8, Topt = 25	Neitsch *et al.* (2011)
WOFOST	Assimilation rate	Multilinear	Tdaytime	Tcrt = c(0, 9, 16, 18, 20, 30, 36, 42)	Supit *et al.* (1994)

*Topt: optimum temperature above or below which stress will occur; a nonstress plateau is assume for curves with two optimum temperatures (e.g., Topt1 and Topt2).
†Tbase: base temperature below which full stress is assumed.
‡Tlim: limiting temperature threshold at which full heat stress is reached.
§Tcrt: critical temperature threshold at which heat stress starts.

reused the APSIM simulation configuration and parameters from the well-calibrated site in Boone, IA (Archontoulis *et al.*, 2014b). We again used meteorological inputs from the Daymet dataset.

Ensemble simulations

The algorithm ensemble for each site consisted of 30 simulation runs (i.e., 10 simulations of heat stress algorithms for different processes × 3 varieties of drought stress algorithms). For heat stress, we constructed ten simulations (SM) that covered (i) two vapor pressure deficit (VPD) calculation methods, (ii) four different representations of heat stress on biomass production, (iii) two heat stress modifiers on grain filling, (iv) three harvest index (HI) models, and (v) one leaf-level photosynthesis model (Fig. 1). Specifically, SM1 is the reference simulation that used the default APSIM algorithms of heat stress on photosynthesis, grain number development, and grain filling. SM2 replaced the default APSIM VPD algorithm, which is purely based on maximum and minimum daily temperature and is hence occasionally criticized for overestimating drought stresses during hot days (Basso &

Table 2 Summary of drought stress algorithm on maize photosynthesis, grain set/fillings, and harvest index. Detailed descriptions are given in Data S3

Model	Process	Conceptual	Function type	References
AgroIBIS	Photosynthesis rate (Vmax)	SWC	Exponential	Kucharik & Brye (2003)
APSIM-Maize	RUE	Ws/Wd	Linear	Keating *et al.* (2003)
	Grain filling	Ws/Wd	Linear	
AquaCrop	Stomatal closure	SWC	Convex curve	Raes *et al.* (2009)
	Harvest index	Complex subroutines		Raes *et al.* (2009)
CERES-Maize	RUE	AT/PT	Linear	López-Cedrón *et al.* (2005)
	Grain filling	AT/PT	Quadratic	López-Cedrón *et al.* (2008)
CropSyst	Water dependent growth	Transpiration efficiency	Linear	Stockle *et al.* (2014)
	Harvest index	Stage-dependent average water stress		
DayCent	GPP	Available water to PET	Linear	Parton *et al.* (1998)
	Carbon allocation	Soil water content	Empirical	
EPIC	RUE	Wu/PT	Linear	Sharpley & Williams (1990)
	Harvest index	Wu/PT	Convex curve	Challinor *et al.* (2004)
GLAM	Transpiration efficiency	Transpiration efficiency		
HYBRID-maize	Assimilation rate	AT/PT	Linear	Yang *et al.* (2013)
CSM-IXIM	Carbon allocation	AT/PT	Exponential	Lizaso *et al.* (2011)
MAIZSIM	Stomatal conductance	Leaf water potential	Logistic	Yang *et al.* (2009)
	Carbon allocation	SWC	Linear	Acock *et al.* (1982)
MONICA	Assimilation	AT/PT	Linear	Sage & Kubien (2007)
PEGASUS	LUE	SWC	Exponential	Deryng *et al.* (2011)
STICS	RUE	SWC	Linear	Brisson *et al.* (2009)
SWAT	RUE	AT/PT	Linear	Neitsch *et al.* (2011)
	Harvest index	AET/PET	Linear	
WOFOST	Assimilation rate	AT/PT	Linear	Supit *et al.* (1994)

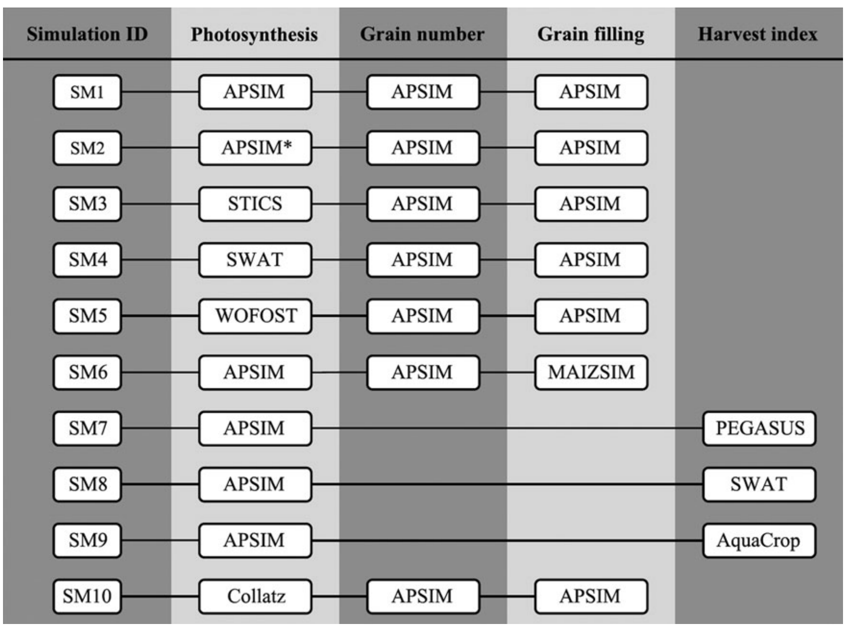

Fig. 1 Framework for using ensemble simulations to compare algorithms at the process level. Heat stress algorithms for each process (i.e., photosynthesis, grain number development, grain-filling rate, and harvest index increment) are listed as bricks. The combination of different bricks for all processes evaluated leads to a simulation (SM).

Ritchie, 2014), with the more common method that requires either daily dew point temperature or relative humidity as input data (Abtew & Melesse, 2013).

SM3, SM4, and SM5 replace the APSIM multilinear temperature stress function on the radiation use efficiency (RUE) with its counterpart in STICS, SWAT, and WOFOST, respectively. It should be noted that STICS uses canopy temperature, which can be calculated by an empirical relation model, instead of air temperature to force the stress function (Data S3.16). SM6 was a simulation using the algorithm of high temperature effect on grain filling from MAIZSIM. SM7, SM8, and SM9 retained the APSIM photosynthesis and biomass production routines, but estimated yield based on the simulation of HI instead of the original grain number × grain-filling rate method. SM7 incorporated the PEGASUS HI method (also used in CropSyst and GLAM), in which potential HI can be reduced due to heat stress around the silking–anthesis stage (i.e., flowering stage). SM8 used the SWAT HI method, which first develops potential HI according to the accumulation of daily heat units, and then calculates the actual HI based on the average water deficit over the growing season. SM9 adopted the HI model from AquaCrop, in which the potential HI can be adjusted either upward or downward by a number of environmental stress factors (Raes *et al.*, 2009). To compare the performance of RUE-based biomass production models with the more mechanistic model of leaf-level CO_2 assimilation processes, we incorporated the coupled photosynthesis–stomatal conductance model for C4 plants according to Collatz *et al.* (1992) as SM10 (Data S6). Similar leaf-level photosynthesis models have been implemented in more recently developed crop models (e.g., AgroIBIS, CSM-IXIM, MAIZSIM, and MONICA). As SM1–SM10 are not fully orthogonal, results from these simulations should not all be compared against each other. The effect of changing the APSIM default VPD algorithm can be observed by comparing SM1 vs. SM2. If the focus is on different parameterizations of heat stress on biomass production, then compare SM1 vs. SM3, SM4, and SM5. Comparing SM1 and SM6 illustrates the difference between two heat stress functions on grain filling. Different implementations of HI algorithms can be evaluated by looking at SM7, SM8, and SM9, while the difference between grain filling vs. the HI method can be compared by looking at the group of SM1 and SM3-5 vs. the group of SM7-9. The effect of replacing an RUE model with leaf-level photosynthesis can be seen by comparing results from SM1 and SM10. On top of each simulation with a particular heat stress algorithm, we further nested three varieties of drought stress algorithms that describe water deficit as a function of SWC, Ws/Wd, or AT/PT. More detailed theoretical backgrounds for each of these algorithms are given in Data S3. Simulations of maize phenology, soil moisture, temperature, and nutrient dynamics were still carried out by the default APSIM platform.

We used Daymet meteorology variables, as mentioned above, to run APSIM. Soil parameters, such as layered soil hydraulic properties and soil organic matter fractions, were extracted from the SSURGO database (Web Soil Survey: http://websoilsurvey.sc.egov.usda.gov). A detailed description for each of these soil parameters is presented in Archontoulis *et al.* (2014a,b). When a farm had several soil types according to SSURGO, we simply selected the one that accounted for the largest fraction, to reduce the computational cost. As a result, we derived Flanagan silt loam soil for the Illinois farm, Chalmers silt clay loam soil for the Indiana farm, and Webster clay soil for the Iowa farm. Management history is critical for models to reproduce the historical trend in maize yield. In rainfed fields, the required management information includes as follows: (i) sowing date, seeding rates and cultivar; (ii) fertilizer type, amount, and timing. We derived most of the information from the NASS survey report, with state-specific details provided in Table S1.

Analysis

To evaluate the APSIM-Maize performance on predicting yield, we calculated the coefficient of determination (R^2) and root mean square error (RMSE) based on simulated yields and the NASS county-level rainfed maize yield data (e.g., Tippecanoe County for the farm from West Lafayette, IN). We further detrended the yield over time by applying a linear regression and then used residuals to calculate the Spearman correlation.

To quantitatively understand the sensitivity of model-simulated biomass and/or yield to heat and drought stress, we further calculated the relative contributions of each stress over the historical period of 1980–2013 and in two future climate scenarios. Simulations were conducted by the standard APSIM-Maize (i.e., SM1) for the Indiana farm. The APSIM framework allowed us to switch on and off a certain stress by setting the corresponding stress function equal to 1 (Data S3.2). The sensitivity of biomass reduction (%) to drought was calculated as:

$$S_{\mathrm{Drought}} = \frac{\left(B_{\mathrm{Drought}} - B_{\mathrm{Potential}}\right)}{B_{\mathrm{Potential}}} \times 100\% \qquad (1)$$

where $B_{\mathrm{Potential}}$ is the simulated biomass from SM1 when stresses that directly limit photosynthesis and grain development are excluded, and B_{Drought} is the value from the simulation that includes drought stress. Likewise, we calculated the sensitivity of biomass accumulation to high temperature as:

$$S_{H\_\mathrm{RUE}} = \frac{\left(B_{\mathrm{Temperature}} - B_{\mathrm{Potential}}\right)}{B_{\mathrm{Potential}}} \times 100\% \qquad (2)$$

in which $B_{\mathrm{Temperature}}$ is the value from the simulation that only applied the temperature response curve to the RUE. The sensitivity of grain growth, and hence yield, to extreme heat was quantified as:

$$S_{H\_\mathrm{Grain}} = \frac{\left(Y_{\mathrm{Heat}} - Y_{\mathrm{potential}}\right)}{Y_{\mathrm{Potential}}} \times 100\% \qquad (3)$$

where $Y_{\mathrm{potential}}$ is the potential yield that considered stresses on biomass accumulation but not heat stress on grain set and grain fill, and Y_{Heat} was the actual yield. To run APSIM-Maize under a projected future climate, we used daily projections from 2006 to 2099 provided by The NASA Earth Exchange Global Daily Downscaled Projections (NEX-GDDP). This downscaled dataset in a spatial resolution of 0.25 degrees was derived from the general circulation models (GCMs) participating in the Coupled Model Intercomparison Project Phase 5 (CMIP5) under two of the four representative concentration pathways (RCPs). The effect of elevated CO_2 on maize growth was not simulated here, as it is beyond the scope of this study and the magnitude of maize yield response to CO_2 is controversial (Leakey *et al.*, 2009). To reduce the computational cost, we selected projections for RCP4.5 and RCP8.5 from 8 representative GCMs (Table S2). The simulations conducted here were enough to extend the quantification of relative contributions of heat and drought stress into the future.

Results

Screening of heat stress functions

Temperature response curves of maize carbon assimilation differ markedly among selected crop models (Fig. 2a). While some models use a single optimum temperature in their response curve (e.g., AgroIBIS and DayCent), others specify a wider range of temperatures (i.e., a plateau) for optimum or near optimum growth. AgroIBIS, EPIC, and SWAT specify 25 °C as the optimal temperature for maize, beyond which heat stress starts to reduce photosynthesis. APSIM, DayCent, and WOFOST use approximately 30 °C as the maximum optimal temperature. CERES and MAIZSIM, using daylight and hourly temperature as the forcing data, have even higher maximum optimal temperature of 33 °C and 32 °C, respectively. The upper limit temperature at which stress reaches its maximum differs substantially among models (Fig. 2a). These differences are also reflected by the observed temperature responses of GPP to APAR ratio (as an approximation of RUE) (Fig. S1). The optimal temperature range for hourly GPP/APAR is roughly 20–31 °C, and the response curve is more like a piecewise linear function. For daylight GPP/APAR, the optimal temperature range is roughly 28–31 °C; this is probably why our simulations produce similar results when using daylight and hourly temperature to drive the algorithms. The optimal temperature for the daily mean GPP/APAR occurs around 25 °C (which agrees with Agro-IBIS, EPIC, and SWAT), and the response curve is more like a quadratic function.

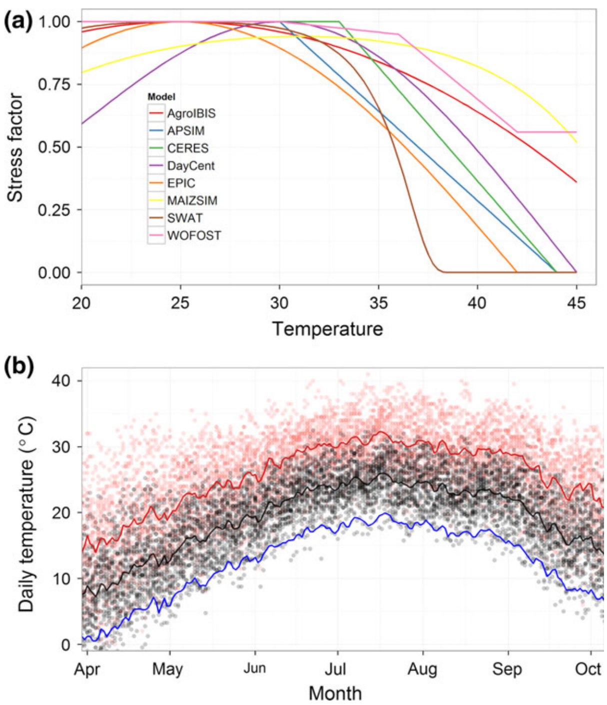

Fig. 2 (a) Temperature response curves used in representative crop models. (b) 34-year (1980–2013) averaged growing season daily maximum (red line), mean (black), and minimum (blue) temperature for the Indiana farm in this study. Red and black dots are daily maximum and mean temperatures for all years, respectively.

The predicted growing season average reduction in photosynthesis due to heat stress did not exceed 2% for most algorithms when forced by daily mean temperature, even in the years of 1988 and 2012, in which severe heat waves were recorded (Fig. 3). When heat stress is simulated using daylight temperature instead of mean daily temperature, yields simulated using all of the algorithms vary interannually with the heat stress factors and become negatively correlated with EDD or KDD (Fig. 3). Algorithms from APSIM, DayCent, EPIC, MAIZSIM, and WOFOST have very high correlations ($r < -0.95$), followed by AgroIBIS ($r = -0.87$). The magnitude of reduction due to heat stress typically remained <5%, except for the EPIC simulation, which decreased by up to 10%. We also tested the effect of increasing simulation time frequency, in which the daily stress is calculated by averaging the every 3 hours prediction and obtained results very close to simulations that use daylight temperature (not shown).

Screening of drought stress functions

During the moist year of 2010 (May–August precipitation was 878 mm), algorithms that calculate stress factor as a function of SWC or Ws/Wd (SWC method and Ws/Wd method hereafter) predicted almost no drought stress, while the algorithm based on AT/PT (AT/PT method) predicted substantial stress over the growing season (Fig. 4a). During the dry year of 2012 (May–August precipitation was 301 mm), all three methods indicated severe drought during the summer, although the magnitude of water shortage predicted by the Ws/Wd method was much greater than the other two methods (Fig. 4b). The more severe drought predicted by Ws/Wd starting in July was likely caused by both the steady decrease

  doi: 10.1111/gcb.13376

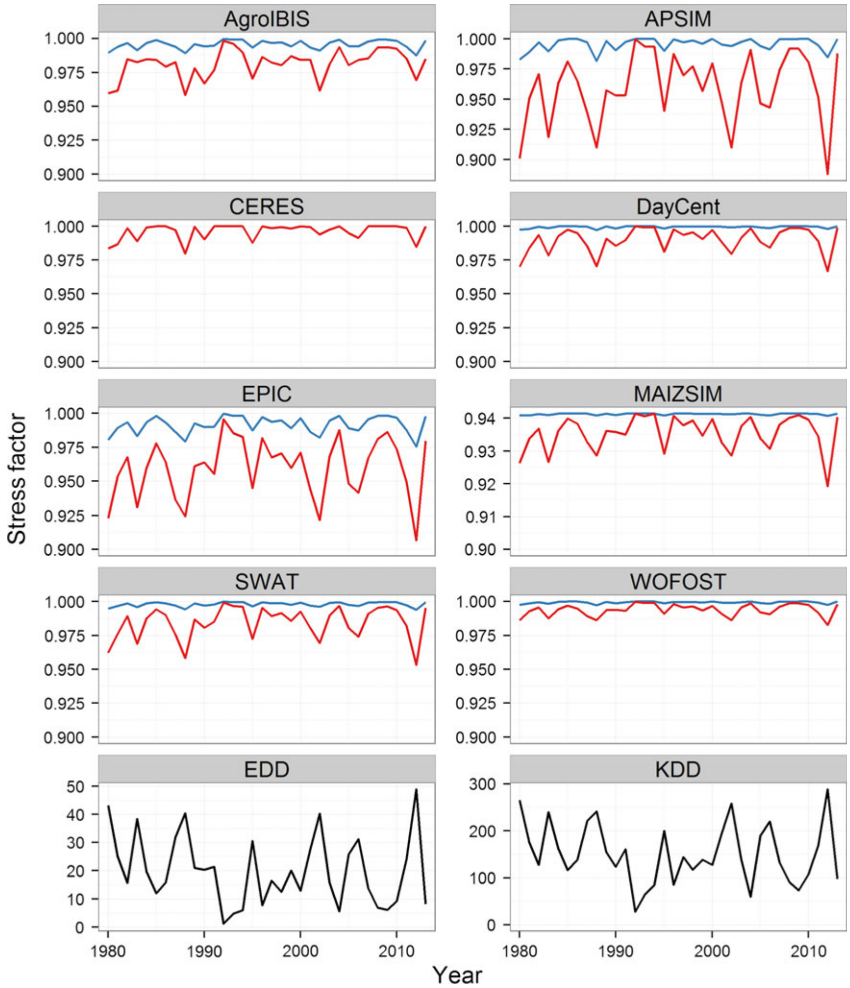

Fig. 3 Effect of the temperature forcing data of algorithms on their predictions of mean annual heat stress (1 for no stress and 0 for full stress) for the Indiana farm. Simulations using daily mean temperature are shown as blue lines, and simulations with daylight temperature are shown as red lines. Note that AquaCrop's algorithm is on a different scale than those from the other models. Indexes of excessive heat, namely extreme degree days (EDD) and killing degree days (KDD) (Data S4), are given for reference.

in soil water supply and the persistent high transpiration demand (Fig. S2). The AT/PT method indicated occasional water deficit in the early growing season, while the other two methods were unresponsive (Fig. 4b). Mean annual drought stress varied substantially across years, fluctuating between 0.7 and 1.0 for years 1980–2013 (Fig. 4c). The stress calculated by the SWC method closely resembled results from the Ws/Wd method (R^2 = 0.9), whereas the AT/PT method differed (R^2 = 0.53 with the SWC method and R^2 = 0.67 with the Ws/Wd method), consistently predicting more severe drought stress.

Comparison between algorithm ensembles

The ensemble simulations generally captured the interannual yield variability for the years 1980–2013 (Fig. S3), with R^2 varying between 0.39 and 0.67, RMSE ranging from 1.089 to 1.557 t ha^{-1}, and Spearman correlation ranging from 0.2 to 0.6 (Fig. 5). Our simulations suggest there are increasing yield trends of 65–80 kg ha^{-1} yr^{-1} (varying among simulations) over the study period. These trends are lower than the value derived from NASS statistics for Tippecanoe, Indiana (i.e., 122 kg ha^{-1} yr^{-1}), but outperform the simulated results in Lobell *et al.* (2014) for Johnson, Iowa, using APSIM-Maize and Drewniak *et al.* (2013) for the average US maize using CLM-Crop that show almost null or even negative yield trends. The improvement is mainly because we explicitly customized the simulations with yearly management information (e.g., planting date, density, and fertilizer amount) according to the NASS database (Table S1). Interestingly, using different drought stress algorithms had little effect on the model predictability, except that the AT/PT method produced slightly worse performance (e.g., SM2 and

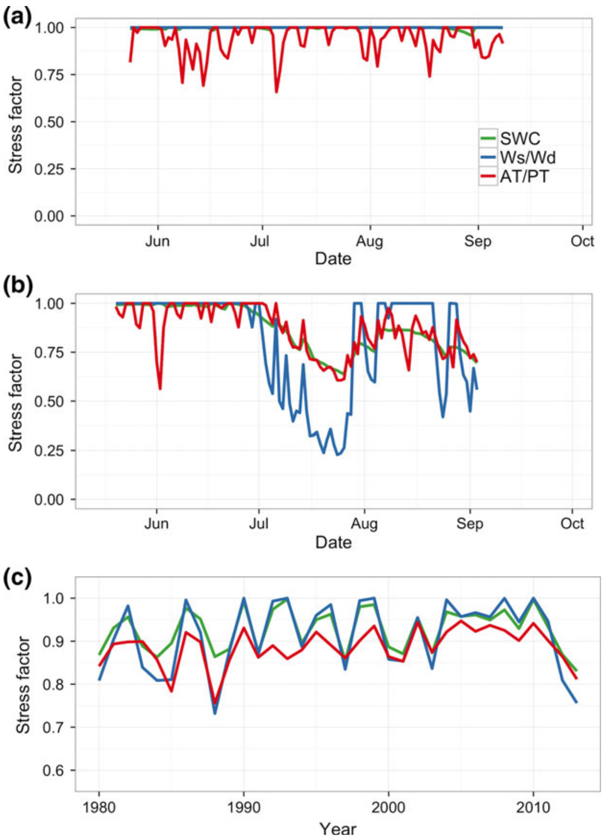

Fig. 4 Drought stress (1 for no stress and 0 for full stress) for the Iowa farm as predicted by different drought stress algorithms. Seasonal dynamics of daily stress factors for the moist year of 2010 (a) and the drought year of 2012 (b). (c) Interannual variability of mean growing season stress factors from 1980 to 2013.

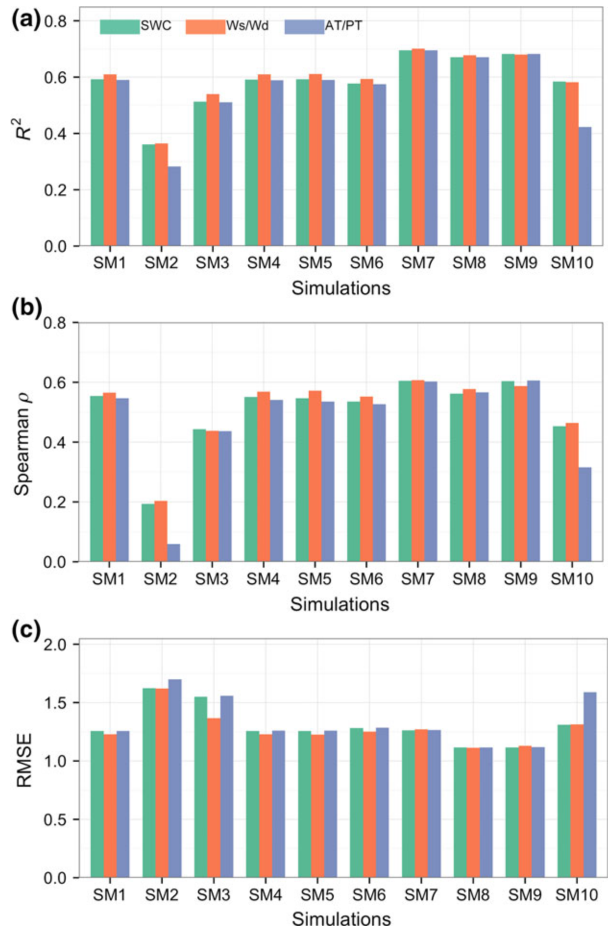

Fig. 5 Evaluation of model performance for the Indiana farm under 30 ensemble simulation (SM) trials (10 heat × 3 drought stress algorithms) with respect to reproducing the USDA county-level yield statistics from 1980 to 2013. Model predictability is measured collectively by (a) R2 and (c) root mean square error (RMSE) derived from the raw data, and (b) Spearman correlation coefficient (ρ) derived from the time detrending data. See Fig. 1 for detailed algorithm combinations for each ensemble.

SM10; Fig. 5a,b). Although mainstream drought stress algorithms produced quite different predictions for the seasonal pattern of water deficit (Fig. 4), they displayed similar capability to represent drought on an annual basis.

Simulations from SM2, with the updated VPD algorithm, generally gave the worst model predictions (smallest R^2 and Spearman correlation, largest RMSE; Fig. 5), but outperformed all other simulations for the extreme drought year of 2012. Other simulations with the default VPD algorithm substantially underestimated yield by as much as 2.9 t ha^{-1} in that year (Fig. S3). Such systematic biases could be a result of the overestimation of VPD and hence crop water demand. In the current version of APSIM, the daily water-limited dry matter production, calculated as soil water supply × transpiration efficiency (TE), is inversely proportional to VPD (Data S3). The overestimation of VPD may lead to unrealistically high water demand and thus greatly overstates water deficits on exceptionally hot days (Basso & Ritchie, 2014). On the other hand, underestimating soil water supply when high VPD continuously depletes soil water could also overestimate the drought stress. Take the extreme dry year of 2012 as an example: weekly maximum VPD was almost 1.1 kPa higher when simulated by the default method than with the conventional method (Fig. S6), which lowered TE and reduced biomass, as water supply was coincidently also exceptionally low. However, because the APSIM-Maize model has long been calibrated with the default VPD calculation route, simply changing the VPD algorithm will not guarantee an improvement in the overall model performance.

Using canopy temperature (SM3) instead of daily mean temperature to calculate heat stress lowered model performance at farms from Indiana (Fig. 5) and Illinois (Fig. S4) and slightly improved model predictions for the Iowa farm (Fig. S5), possibly because the empirical canopy temperature model we adapted from STICS is only valid under a limited set of conditions. The simulated mean daily canopy temperature was generally higher than the air temperature measured at 2 m height, but mostly no more than 3°C (Fig. S7), whereas the difference observed in rainfed fields ranged from −2 to 7.5°C (Siebert *et al.*, 2014). Switching between heat stress algorithms made little difference for predicted yield variability (i.e., SM1 vs. SM4-6), confirming that current crop models are insensitive to heat stress. Although it is difficult to recommend any algorithm over the others under contemporary climate conditions, crop modelers should keep in mind that these algorithms may diverge substantially when being used for future projections.

Simulations with the HI method consistently outperformed the others in terms of capturing the yield variability ($R^2 > 0.64$) and minimizing the prediction error (Fig. 5). SM8 and SM9 performed slightly better than SM7, which used the PEGASUS algorithm, possibly because PEGASUS does not include water stress like the former two algorithms, but only considers heat stress around the silking–anthesis period when calculating the actual HI (Deryng *et al.*, 2014). Potential HI for AquaCrop can be more conservative (e.g., 0.5 in this study), because AquaCrop has incorporated a mechanism through which crops generally produce excessive flowers to help recover once environmental constraints on pollination are ameliorated (Raes *et al.*, 2009; Data S3). The parameter of potential HI for SWAT should be set slightly higher than for the other two models to obtain acceptable results (e.g., potential HI = 0.55 in this study), as the HI in SWAT is often stressed more than that in the PEGASUS model and will not be compensated by additional flowers as in AquaCrop.

Last but not least, the leaf-level photosynthesis algorithm had a similar prediction bias (RMSE = 1.272 t ha^{-1}) and yield variability ($R^2 = 0.54$) as the RUE-based simulation (SM1 vs. SM10; Fig. 5), despite its more complex model structure and heavier computational load (if solving coupled equations uses a numerical iteration method). It should be noted that the Collatz model does not explicitly consider N limitation when calculating the gross CO_2 assimilation (Collatz *et al.*, 1992) and is thus less responsive to the historical increase in fertilizer applications (Fig. S3).

Past and projected future contributions of heat and drought stress to yield loss

Yield losses at the Indiana farm due to climatic stress were attributed more to water deficits than suboptimal temperatures (hot or cold; Fig. 6), and thus, the losses caused by excess heat were even smaller. The direct losses from higher than optimal temperature were mostly trivial and accounted for no more than 6% even in the notoriously hot years of 1988 and 2012, while the losses from water stress were more than 10% in several years and peaked at 30% in 2012. However, part of the water stress impact could be an indirect effect of high temperature, as warming increases water demand via elevating the VPD and at the same time decreases soil water storage by accelerating transpiration over short time periods (Lobell *et al.*, 2013).

Under projected future climates, the models suggest drought will continue to play a critical role in reducing the maize production at the Indiana farm, and the stress will intensify faster under the high emission scenario (Fig. 7). Average biomass reduction due to drought will increase from 15% in the 2000s to 20% and 27% at the 2090s under RCP4.5 and RCP8.5 scenarios, respectively. The influence of high temperature on biomass accumulation is predicted to be small under RCP4.5, but becomes increasingly prominent after 2050s under RCP8.5. In a few years warmer climates increase yields, possibly because the positive effect of moderate warming on the rate of grain filling overcomes the negative effect on other processes. Extreme heat only occasionally damages simulated maize production in the first half of the 21st century, but reduces grain number and yield with greater frequency and intensity after the 2050s, especially under the RCP8.5 scenario (Fig. 7c). It should be noted, however, the relative importance of drought vs. heat is specific to the US Midwest and may differ in more humid regions such as Europe.

Discussion

Lessons from the review of algorithms

Heat stress functions can be effective when based on T_{mean}, daylight, or hourly temperature as long as they

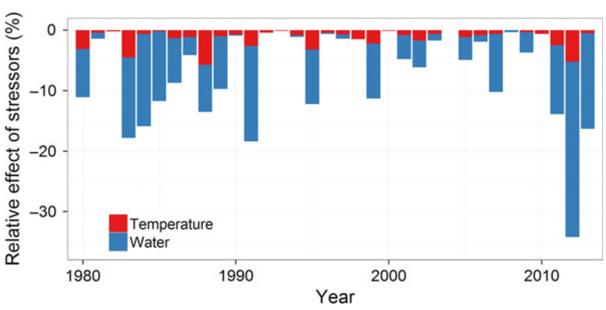

Fig. 6 Percentage yield reduction attributed to temperature and water stress on the Indiana farm from 1980 to 2013, as simulated using the standard APSIM-Maize model.

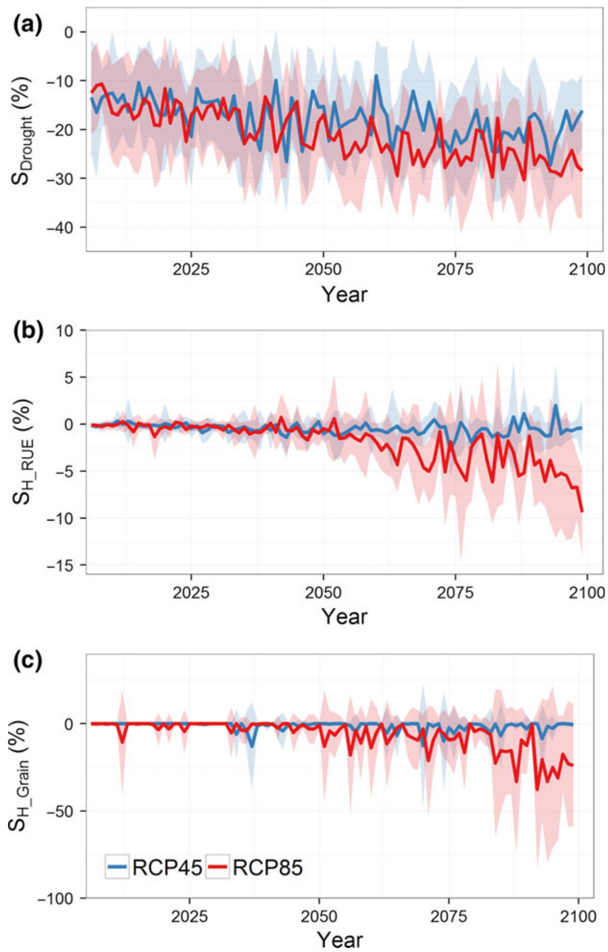

Fig. 7 The effects of drought (a), high temperature via photosynthesis (b), and heat via grain development (c) on maize yield for the Indiana farm under two Representative Concentration Pathway (RCP) scenarios. Solid lines are mean predictions from eight general circulation models (GCMs), and shaded areas represent one standard deviation.

are parameterized correctly. However, it is very likely that a few models that base their temperature responses of RUE on T_{mean} actually have functions that were parameterized based on an hourly (or instantaneous) temperature response. For crop models that use daily T_{mean} to calculate heat stress factors, the optimal temperature threshold for algorithms should be smaller than algorithms using daylight or hourly temperature. The likely maximum optimal temperature for a T_{mean} function is around 25°C, which is smaller than the critical temperature threshold for maize growth (i.e., ~30°C) derived from large-scale statistics by Schlenker & Roberts (2009) and Lobell *et al.* (2013). Nonetheless, the literature-suggested temperature threshold is very close to the maximum optimal daylight or hourly temperature for RUE of 31–32°C. Interestingly, as we move from regional scale models (e.g., AgroIBIS, EPIC, and SWAT) to cropping systems models (e.g., WOFOST) and then plant level models (e.g., CERES, and MAIZSIM), the T_{opt} increases, indicating the need to consider different T_{opt} for different scales of simulation analysis (region, crop, leaf-level). One follow-up concern is that these temperature thresholds may vary across space, given that the cultivars planted could be different from one place to another as a result of years of breeding and selection. While the spatial pattern of an optimal temperature threshold deserves further investigation, we also suggest that crop modelers consider replacing this type of hard-coded temperature threshold with uncertain parameters, to increase model agility (Mendoza *et al.*, 2015).

The use of daylight temperature instead of instead of T_{mean} improves model performance by making heat stress algorithms responsive, likely because the current parameterizations of heat stress algorithms in most crop models that use daily mean temperature happen to be close to the RUE response curve to daylight temperature (Figs 2a and S1). This simple modification is very easy to implement and is further justified when the difference between 3-hour simulations and the use of $\frac{T_{max}+T_{mean}}{2}$ is very small on either a daily or an annual basis. Shortening the simulation time step certainly works because it allows the algorithm to reproduce the diurnal cycle of air temperature and hit those time points when temperature is significantly higher than the threshold. To control the computational cost that includes major re-parameterization, crop modelers would not have to run the whole model with higher time frequency, but could simply run the subroutine used to calculate stress factors.

The behaviors of drought stress algorithms were close to our expectations. In general, predictions made by the SWC method were less severe but smoother, possibly because the use of a multilayer tipping-bucket model in the APSIM. As maize roots can normally penetrate to 1.5-2 meters depth and withdraw water throughout the whole soil profile (Hochholdinger & Tuberosa, 2009), crop models often calculate water stress by averaging stress factors across all of the layers. However, simulated soil moisture of deep layers in many crop models normally had very small fluctuations, therefore minimizing simulated water stress for the whole soil column. The AT/PT method, which calculates potential transpiration with the Priestley-Taylor equation (Priestley & Taylor, 1972), showed substantial daily fluctuation, and tended to overestimate drought stress when there was no or mild soil water shortage. Sau *et al.* (2004) also reported that the use of Priestley-Taylor equation tends to overpredict potential ET measured under irrigated and rainfed conditions in southern Spain, which reduces stress factors when AT is fixed, and therefore underestimates LAI, biomass, and grain yield. The use of the FAO56 ET method (Allen *et al.*, 1998) has been shown to perform better than the Priestley-Taylor method (Saseendran *et al.*, 2008), but requires more detailed ground observational data as input which may not be widely available (e.g., wind velocity and relative humidity). However, even if the calculation of PT can capture daily weather fluctuations well, how fast crops can respond to those fluctuations remains an open question. The Ws/Wd method, which is based on the concept of transpiration efficiency (Data S3.2), predicted little water stress during the cool early growing season, likely because Wd is small as a result of: (i) low VPD at low temperature and hence high TE; (ii) low dry matter accumulation rates given the low temperatures and less radiation interception in the early season. During the drought year of 2012, the Ws/Wd method predicted substantially more severe drought than the AT/PT method due to both high Wd values and low Ws (Fig. S2). It should also be noted that a recent conceptual theoretical analysis (Basso & Ritchie, 2014) argued that APSIM tends to overestimate VPD during hot summers.

Lessons from the ensemble simulation

The consistent underestimation of the yield increase trend by all simulations may be a consequence of simulating a single cultivar for the whole study period and in all of the different locations (Fig. 5a). It is well established that farmers change cultivars very frequently, and cultivars vary substantially in their yield potential as a result of differences in traits such as relative maturity (Kumudini *et al.*, 2014), light use efficiency (Tollenaar & Aguilera, 1992; Singer *et al.*, 2011), and genetically modified stress-tolerance (Xu *et al.*, 2013). While such cultivar information is more difficult to obtain, crop modelers can inversely estimate spatiotemporal variations of cultivar-specific parameters against in-situ measurements. Given the very limited number of existing case studies (Sakamoto *et al.*, 2010; Jones *et al.*, 2011; Archontoulis *et al.*, 2014a,b), this area deserves more research effort in the future.

Contrary to our expectations, the seemingly simple HI method outperformed more mechanistic methods that account for grain numbers and grain filling. A possible explanation is that the HI method has been parameterized based on historical county-level yield statistics data that was used to evaluate models performance here. Moreover, when simulating maize yield with more mechanistic algorithms, climate variability has already been largely represented in the biomass estimates, so that additional steps to simulate grain number and grain filling based on the concept of carbon source and sink lead to a greater uncertainty than obtained with the HI method. On the other hand, models that explicitly simulate kernel development can provide estimates of grain number, sugar, and oil content, all of which are commercially valuable information (Borrás *et al.*, 2002). In short, more complex and mechanistic algorithms are not necessarily better than simpler alternatives. The pros and cons of simple algorithms largely depend on the model application scale and variable of interest.

Although the leaf-scale photosynthesis model showed no apparent advantages in terms of predicting yield, it should be considered as a research frontier for next generation model development (Boote *et al.*, 2013). The conventional RUE-based crop models have hit a bottleneck, in that they lack leaf-level physiological processes, and hence cannot disentangle interactions between photosynthesis and many well-known regulating factors such as light, CO_2, leaf energy, leaf water, and enzyme status (Lizaso *et al.*, 2005). For example, elevated atmospheric CO_2 is believed to mitigate water stress in maize by reducing stomatal conductance and improving water use efficiency (Leakey *et al.*, 2006; Hussain *et al.*, 2013), but how much this will truly benefit yield is open to debate (Leakey *et al.*, 2009; Boote *et al.*, 2013; Urban *et al.*, 2015). In fact, a negative feedback exists between improved water use efficiency and canopy temperature and VPD, because lower transpiration will reduce latent heat flux from canopy to the atmosphere, causing foliage temperatures to rise, which could again increase transpiration (Lobell *et al.*, 2013). Improved crop modeling at the leaf scale that couples CO_2, water, and energy is thus needed.

Reflections on future crop model improvements

Overall, our analysis shows that algorithms from representative maize models do not adequately capture the impact of climate extremes on maize photosynthesis and yield. These conclusions are consistent with several other model comparison studies for cereal crops under various growth conditions (e.g., Asseng *et al.*, 2013; Eitzinger *et al.*, 2013; Bassu *et al.*, 2014). Knowledge gaps and promising research frontiers for improving the predictability and credibility of current crop models have been discussed in a number of review papers (Boote *et al.*, 2013; Parent & Tardieu, 2014; Barlow *et al.*, 2015; Rezae *et al.*, 2015). Based on our analyses, we highlight the following three features that have not been well addressed in existing crop models.

First, crop models need better mechanisms to handle climate and weather extremes. Existing temperature and moisture response functions of many physiological processes used by crop models to capture the climate variability are mainly summaries of observed historical statistics (Reichstein *et al*., 2013) and hence are questionable when used to fit novel climate conditions. For instance, the extremely high yield reduction predicted by the standard APSIM in the 2090s should be treated with caution, as it has not been validated at those novel bioclimatic scenarios. Regarding time scale, heat waves may happen very quickly—within a window of a few hours—and therefore is beyond the current simulation capacity of most crop models. CropSyst has recently incorporated a mechanism to discount biomass production when high temperatures last for more than 4 hours (Alderman *et al*., 2014). In addition, a perspective from ecosystem modeling suggests defining extreme climatic events as ‘an episode or occurrence in which a statistically rare or unusual climatic period alters ecosystem structure’ (Smith, 2011). In this sense, crop models should go beyond the current continuous reduction functions and incorporate mechanisms to capture heat and drought stress that occurs singly, coincidently or when one follows another, and whose impact may or may not be reversible. Existing models only have very limited implementations for events-based simulation. For example, in APSIM-Maize high temperatures immediately following emergence will kill a fraction of plants. The implementation of a response of grain number set to heat extremes in APSIM and DSSAT is an early attempt to account for the carryover effect, although its parameterization is not adequately reliable due to limited experimental data. Other models, including CropSyst, GLAM, MONICA, and PEGAUS, implement a reduction in HI when there is heat stress around the flowering stage.

Second, although the importance of considering canopy temperature in quantifying the heat stress impact has been emphasized quite often in recent years (Siebert *et al*., 2014; Rezae *et al*., 2015), potential losses from increasing nighttime temperature also deserve adequate attention. Nighttime warming has been shown to negatively affect plant growth across the Northern Hemisphere, because it boosts nighttime plant respiration that consumes carbon accumulated during daylight photosynthesis (Peng *et al*., 2013). Evidence also suggests that damage from nighttime heat stress is amplified during the reproductive phases and that nighttime warming was partly responsible for the lower productivity and reduced kernel quality observed across the US Corn Belt in 2010 and 2012 (Hatfield *et al*., 2014). With the number of hot nights projected to increase by as much as 30%, yield reductions will become more prevalent (Hatfield *et al*., 2011). However, none of the models we reviewed explicitly considered the direct impact of nighttime warming. Crop models with leaf-level photosynthesis algorithms can be easily adapted to account for nighttime heat (e.g., AgroIBIS, CSM-IXIM, and MAIZSIM), although they have not been well parameterized and tested. MONICA also uses a mechanistic photosynthesis model, but its daily time step certainly obscured the signal of high nighttime temperature (Supplementary material). For models using the RUE approach, the nighttime temperature effect could be considered by incorporating a new limiting factor as a function of nighttime temperature when calculating the daily biomass accumulation, or by adding a reduction term elsewhere (e.g., when allocating the dry matter to grains).

Finally, the best way to coordinate multiple stresses needs further investigation. For those RUE-based models, the minimum of heat and drought stress factors is normally used to limit potential biomass production (e.g., APSIM-Maize, CropSyst, CSM-CERES, and SWAT), while a product of both is applied in PEGASUS and STICS. In some cases, VPD is further used to adjust the potential RUE or TE (e.g., APSIM, CropSyst, SWAT, and GLAM). For leaf-level photosynthesis models, the temperature effect is supposed to be captured by the temperature dependency of each parameter, and water stress is reflected in the stomatal conductance. But AgroIBIS also adjusts maximum photosynthetic rate by a water stress factor, and MAIZSIM limits stomatal conductance by a function of leaf water potential. This variety of approaches begs the question: Do any or all of these forms lead to double accounting of heat and drought stresses? To our knowledge, no studies have answered this question. When simulating yield formation, either via grain development or the HI method, some models purely use heat or drought stress alone and some models use both (Tables 1 and 2). Given that these crop models are individually developed and their main purpose is to predict biomass or yield variability, the inconsistency in the organization of these stress factors is quite understandable. However, this question should be answered because: (i) current models may give the right result but for the ‘wrong reasons’, that is, despite being based on questionable algorithms, and (ii) the lack of an answer hinders the assimilation of newly discovered stress mechanisms. One possible solution for mechanistic models is to compare intermediate model outputs (such as LAI, canopy level assimilation) to intermediate measurements (Boote *et al*., 2013), while for RUE-based models more efforts are needed.

In short, our study identifies the model formulations that best predict the impacts of heat and drought stress on maize biomass production and yield and recognizes gaps to further reduce the prediction uncertainty. The framework presented here can be applied to modeling other crop physiological processes and factors (e.g., phenology, chill, and canopy transpiration) and used to improve yield predictions of other crops in a wide variety of crop models, thus is a significant advance in the crop modeling research.

Acknowledgements

We thank Graeme Hammer for helpful comments on this manuscript. We acknowledge the Information Technology at Purdue Research Computing (RCAC) for computing support. This research was funded to Q.Z. through a NSF Project (Grant IIS-1028291; A Paradigm Shift in Ecosystem and Environmental Modeling: An Integrated Stochastic, Deterministic, and Machine Learning Approach).

References

Abtew W, Melesse A (2013) Vapor pressure calculation methods. In: *Evaporation and Evapotranspiration* (eds Abtew W, Melesse A), pp. 53–62. Springer, Dordrecht.

Acock B, Reddy VR, Whisler FD, Baker DN, McKinion JM, Hodges HF, Boote KJ (1982) *The Soybean Crop Simulator GLYCIM: Model Documentation*. U.S. Gov. Print. Office, Washington, DC. PB85171163/AS.

Alderman P, Quilligan E, Asseng S, Ewert F, Reynolds M (2014) Proceedings of the Workshop on Modeling Wheat Response to High Temperature. El Batan, Texcoco, Mexico; 19-21 Jun 2013, CIMMYT.

Allen RG, Pereira LS, Raes D, Smith M (1998). Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. FAO, Rome, 300(9), D05109.

Archontoulis SV, Miguez FE, Moore KJ (2014a) A methodology and an optimization tool to calibrate phenology of short-day species included in the APSIM PLANT model: Application to soybean. *Environmental Modelling & Software*, **62**, 465–477.

Archontoulis SV, Miguez FE, Moore KJ (2014b) Evaluating APSIM Maize, soil water, soil nitrogen, manure, and soil temperature modules in the midwestern United States. *Agronomy Journal*, **106**, 1025–1040.

Asseng S, Ewert F, Rosenzweig C *et al.* (2013) Uncertainty in simulating wheat yields under climate change. *Nature Climate Change*, **3**, 827–832.

Barlow K, Christy B, O'leary G, Riffkin P, Nuttall J (2015) Simulating the impact of extreme heat and frost events on wheat crop production: A review. *Field Crops Research*, **171**, 109–119.

Basso B, Ritchie J (2014) Temperature and drought effects on maize yield. *Nature Climate Change*, **4**, 233–233.

Bassu S, Brisson N, Durand JL *et al.* (2014) How do various maize crop models vary in their responses to climate change factors? *Global Change Biology*, **20**, 2301–2320.

Boote KJ, Jones JW, White JW, Asseng S, Lizaso JI (2013) Putting mechanisms into crop production models. *Plant, Cell & Environment*, **36**, 1658–1672.

Borrás L, Curá JA, Otegui ME (2002) Maize kernel composition and post-flowering source-sink ratio. *Crop Science*, **42**, 781–790.

Brisson N, Launay M, Mary B, Beaudoin N (2009) *Conceptual Basis Formalisations and Parameterization of the STICS Crop Model*. Editions Quae, Versailles.

Bruinsma J (2009) *The Resource Outlook to 2050. By How Much do Land, Water and Crop Yields Need to Increase by 2050?* FAO Expert Meeting on How to Feed the World in 2050. FAO, Rome.

Butler EE, Huybers P (2013) Adaptation of US maize to temperature variations. *Nature Climate Change*, **3**, 68–72.

Carberry P, Muchow R, Mccown R (1989) Testing the CERES-Maize simulation model in a semi-arid tropical environment. *Field Crops Research*, **20**, 297–315.

Challinor A, Wheeler T, Craufurd P, Slingo J, Grimes D (2004) Design and optimization of a large-area process-based model for annual crops. *Agricultural and Forest Meteorology*, **124**, 99–120.

Challinor A, Wheeler T, Craufurd P, Slingo J (2005) Simulation of the impact of high temperature stress on annual crop yields. *Agricultural and Forest Meteorology*, **135**, 180–189.

Challinor A, Wheeler T, Hemming D, Upadhyaya H (2009) Ensemble yield simulations: crop and climate uncertainties, sensitivity to temperature and genotypic adaptation to climate change. *Climate Research*, **38**, 117–127.

Collatz GJ, Ribas-Carbo M, Berry J (1992) Coupled photosynthesis-stomatal conductance model for leaves of C4 plants. *Functional Plant Biology*, **19**, 519–538.

Coumou D, Rahmstorf S (2012) A decade of weather extremes. *Nature Climate Change*, **2**, 491–496.

Deryng D, Sacks W, Barford C, Ramankutty N (2011) Simulating the effects of climate and agricultural management practices on global crop yield. *Global Biogeochemical Cycles*, **25**, GB2006.

Deryng D, Conway D, Ramankutty N, Price J, Warren R (2014) Global crop yield response to extreme heat stress under multiple climate change futures. *Environmental Research Letters*, **9**, 034011.

Donatelli M, Bregaglio S, Confalonieri R, De Mascellis R, Acutis M (2014) A generic framework for evaluating hybrid models by reuse and composition–A case study on soil temperature simulation. *Environmental Modelling & Software*, **62**, 478–486.

Drewniak B, Song J, Prell J, Kotamarthi V, Jacob R (2013) Modeling agriculture in the community land model. *Geoscientific Model Development*, **6**, 495–515.

Eitzinger J, Thaler S, Schmid E *et al.* (2013) Sensitivities of crop models to extreme weather conditions during flowering period demonstrated for maize and winter wheat in Austria. *The Journal of Agricultural Science*, **151**, 813–835.

Grant R (1989) Simulation of carbon assimilation and partitioning in maize. *Agronomy Journal*, **81**, 563–571.

Hammer GL, van Oosterom E, McLean G, Chapman SC, Broad I, Harland P, Muchow RC (2010) Adapting APSIM to model the physiology and genetics of complex adaptive traits in field crops. *Journal of Experimental Botany*, **61**, 2185–2202.

Hatfield JL, Boote KJ, Kimball B *et al.* (2011) Climate impacts on agriculture: implications for crop production. *Agronomy Journal*, **103**, 351–370.

Hatfield J, Takle G, Grotjahn R *et al.* (2014) Chapter 6: Agriculture. In: *Climate Change Impacts in the United States: The Third National Climate Assessment* (eds Melillo JM, Richmond TC, Yohe GW), pp. 150–174. U.S. Global Change Research Program, Washington, DC.

Hochholdinger F, Tuberosa R (2009) Genetic and genomic dissection of maize root development and architecture. *Current Opinion in Plant Biology*, **12**, 172–177.

Holzworth DP, Snow V, Janssen S *et al.* (2015) Agricultural production systems modelling and software: current status and future prospects. *Environmental Modelling & Software*, **72**, 276–286.

Hussain MZ, Vanloocke A, Siebers MH *et al.* (2013) Future carbon dioxide concentration decreases canopy evapotranspiration and soil water depletion by field-grown maize. *Global Change Biology*, **19**, 1572–1584.

IPCC (2012) Summary for policymakers. In: *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation: Special Report of the Intergovernmental Panel on Climate Change* (eds Field CB, Barros V, Stocker TF, Dahe Q), pp. 1–9. Cambridge University Press, Cambridge.

Jones JW, Hoogenboom G, Poter CH *et al.* (2003) The DSSAT cropping system model. *European Journal of Agronomy*, **18**, 35–265.

Jones JW, He J, Boote KJ, Wilkens P, Porter C, Hu Z (2011) Estimating DSSAT cropping system cultivar-specific parameters using Bayesian techniques. In: *Methods of Introducing System Models into Agricultural Research* (eds Ahuja LR, Ma L), pp. 365–394. American Society of Agronomy, Crop Science Society of America, Soil Science Society of America, Madison, WI.

Keating BA, Carberry PS, Hammer GL *et al.* (2003) An overview of APSIM, a model designed for farming systems simulation. *Modelling Cropping Systems: Science, Software and Applications*, **18**, 267–288.

Kucharik CJ, Brye KR (2003) Integrated BIosphere Simulator (IBIS) yield and nitrate loss predictions for Wisconsin maize receiving varied amounts of nitrogen fertilizer. *Journal of Environmental Quality*, **32**, 247–268.

Kumudini S, Andrade F, Boote K *et al.* (2014) Predicting maize phenology: intercomparison of functions for developmental response to temperature. *Agronomy Journal*, **106**, 2087–2097.

Leakey ADB, Uribelarrea M, Ainsworth EA, Naidu SL, Rogers A, Ort DR, Long SP (2006) Photosynthesis, productivity, and yield of maize are not affected by open-air elevation of CO_2 concentration in the absence of drought. *Plant Physiology*, **140**, 779–790.

Leakey AD, Ainsworth EA, Bernacchi CJ, Rogers A, Long SP, Ort DR (2009) Elevated CO_2 effects on plant carbon, nitrogen, and water relations: six important lessons from FACE. *Journal of Experimental Botany*, **60**, 2859–2876.

Lizaso J, Batchelor W, Boote K, Westgate M (2005) Development of a leaf-level canopy assimilation model for CERES-Maize. *Agronomy Journal*, **97**, 722–733.

Lizaso J, Boote KJ, Jones J, Porter C, Echarte L, Westgate M, Sonohat G (2011) CSM-IXIM: a new maize simulation model for DSSAT version 4.5. *Agronomy Journal*, **103**, 766–779.

Lobell D, Hammer GL, McLean G, Messina C, Roberts MJ, Schlenker W (2013) The critical role of extreme heat for maize production in the United States. *Nature Climate Change*, **3**, 497–501.

Lobell D, Roberts MJ, Schlenker W, Braun N, Little B, Rejesus R, Hammer GL (2014) Greater Sensitivity to Drought Accompanies Maize Yield Increase in the U.S. Midwest. *Science*, **344**, 516–519.

López-Cedrón FX, Boote KJ, Ruíz-Nogueira B, Sau F (2005) Testing CERES-Maize versions to estimate maize production in a cool environment. *European Journal of Agronomy*, **23**, 89–102.

López-Cedrón FX, Boote KJ, Piñeiro J, Sau F (2008) Improving the CERES-Maize model ability to simulate water deficit impact on maize production and yield components. *Agronomy Journal*, **100**, 296–307.

Mallya G, Zhao L, Song X, Niyogi D, Govindaraju R (2013) 2012 Midwest drought in the United States. *Journal of Hydrologic Engineering*, **18**, 737–745.

Martre P, Wallach D, Asseng S *et al.* (2015) Multimodel ensembles of wheat growth: many models are better than one. *Global Change Biology*, **21**, 911–925.

Mendoza P, Clark M, Barlage M, Rajagopalan B, Samaniego L, Abramowitz G, Gupta H (2015) Are we unnecessarily constraining the agility of complex process-based models? *Water Resources Research*, **51**, 716–728.

Moriondo M, Giannakopoulos C, Bindi M (2011) Climate change impact assessment: the role of climate extremes in crop yield simulation. *Climatic Change*, **104**, 679–701.

Neitsch SL, Arnold JG, Kiniry JR, Williams JR (2011) *Soil and Water Assessment Tool Theoretical Documentation Version 2009*. Texas Water Resources Institute, College Station, TX.

Ort DR, Long SP (2014) Limits on Yields in the Corn Belt. *Science*, **344**, 484–485.

Parent B, Tardieu F (2014) Can current crop models be used in the phenotyping era for predicting the genetic variability of yield of plants subjected to drought or high temperature? *Journal of Experimental Botany*, **65**, 6179–6189.

Parton WJ, Hartman M, Ojima D, Schimel D (1998) DAYCENT and its land surface submodel: description and testing. *Global and Planetary Change*, **19**, 35–48.

Peng S, Piao S, Ciais P *et al.* (2013) Asymmetric effects of daytime and night-time warming on Northern Hemisphere vegetation. *Nature*, **501**, 88–92.

Priestley CHB, Taylor RJ (1972) On the assessment of surface heat flux and evaporation using large-scale parameters. *Monthly Weather Review*, **100**, 81–82.

Raes D, Steduto P, Hsiao TC, Fereres E (2009) AquaCrop the FAO crop model to simulate yield response to water: II. Main algorithms and software description. *Agronomy Journal*, **101**, 438–447.

Reichstein M, Bahn M, Ciais P *et al.* (2013) Climate extremes and the carbon cycle. *Nature*, **500**, 287–295.

Rezae EE, Webber H, Gaiser T, Naab J, Ewert F (2015) Heat stress in cereals: mechanisms and modelling. *European Journal of Agronomy*, **64**, 98–113.

Rosenzweig C, Jones J, Hatfield J *et al.* (2013) The agricultural model intercomparison and improvement project (AgMIP): protocols and pilot studies. *Agricultural and Forest Meteorology*, **170**, 166–182.

Sage RF, Kubien DS (2007) The temperature response of C3 and C4 photosynthesis. *Plant, Cell & Environment*, **30**, 1086–1106.

Sakamoto T, Wardlow BD, Gitelson AA, Verma SB, Suyker AE, Arkebauer TJ (2010) A two-step filtering approach for detecting maize and soybean phenology with time-series MODIS data. *Remote Sensing of Environment*, **114**, 2146–2159.

Saseendran S, Ahuja L, Ma L, Timlin D, Stöckle C, Boote K, Hoogenboom G (2008) Current water deficit stress simulations in selected agricultural system models. In: *Response of Crops to Limited Water: Understanding and Modeling Water Stress Effects on Plant Growth Processes* (eds Ahuja LR, Reddy VR, Saseendran SA, Yu Q), pp. 1–38. American Society of Agronomy, Crop Science Society of America, Soil Science Society of America, Madison, WI.

Sau F, Boote KJ, Bostick WM, Jones JW, Mínguez MI (2004) Testing and improving evapotranspiration and soil water balance of the DSSAT crop models. *Agronomy Journal*, **96**, 1243–1257.

Schlenker W, Roberts MJ (2009) Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 15594–15598.

Sharpley A, Williams J (1990) EPIC-erosion/productivity impact calculator: 1. Model documentation. Technical Bulletin-United States Department of Agriculture.

Siebert S, Ewert F, Rezaei EE, Kage H, Graß R (2014) Impact of heat stress on crop yield—on the importance of considering canopy temperature. *Environmental Research Letters*, **9**, 044012.

Singer JW, Meek DW, Sauer TJ, Prueger JH, Hatfield JL (2011) Variability of light interception and radiation use efficiency in maize and soybean. *Field Crops Research*, **121**, 147–152.

Smith MD (2011) The ecological role of climate extremes: current understanding and future prospects. *Journal of Ecology*, **99**, 651–655.

Stockle CO, Kemanian AR, Nelson RL, Adam JC, Sommer R (2014) CropSyst model evolution: from field to regional to global scales and from research to decision support systems. *Environmental Modeling and Software*, **62**, 361–369.

Supit I, Hooijer A, Van Diepen C (1994) *System Description of the Wofost 6.0 Crop Simulation Model Implemented in CGMS*. Joint Research Centre; the European Commission, Wageningen.

Tollenaar M, Aguilera A (1992) Radiation use efficiency of an old and a new maize hybrid. *Agronomy Journal*, **84**, 536–541.

Urban DW, Sheffield J, Lobell DB (2015) The impacts of future climate and carbon dioxide changes on the average and variability of US maize yields under two emission scenarios. *Environmental Research Letters*, **10**, 045003.

Xu Z, Hennessy DA, Sardana K, Moschini G (2013) The realized yield effect of genetically engineered crops: US maize and soybean. *Crop Science*, **53**, 735–745.

Yang Y, Kim S-H, Timlin D, Fleisher D, Quebedeaux B, Reddy V (2009) Simulating canopy transpiration and photosynthesis of corn plants under contrasting water regimes using a coupled model. *Transactions of the ASABE*, **52**, 1011–1102.

Yang HS, Dobermann A, Cassman KG, Walters DT, Grassini P (2013) *Hybrid-Maize (ver. 2013.4). A Simulation Model for Corn Growth and Yield*. Nebraska Cooperative Extension, University of Nebraska-Lincoln, Lincoln, NE.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1 Annual management information reported by USDA National Agricultural Statistics Service.
Table S2 General circulation models (GCMs) used in this study.
Figure S1 Temperature response of radiation use efficiency derived from AmeriFlux hourly observations at Mead rainfed maize, Mead, Nebraska.
Figure S2 Supplementary information for the 2012 Iowa drought at Agricultural Engineering and Agronomy Research Farms of Iowa State University, Boone, IA (42.02°, −93.78°).
Figure S3 Time series of APSIM simulated (red lines) maize yield and NASS county level yield statistics (black lines) for the Indiana farm from 1980 to 2013.
Figure S4 Evaluation of model performance for the Illinois farm under 30 ensemble simulation trials (10 heat × 3 drought stress algorithms) with respect to reproducing the USDA county-level yield statistics from 1980 to 2013.
Figure S5 Evaluation of model performance for the Iowa farm under 30 ensemble simulation trials (10 heat × 3 drought stress algorithms) with respect to reproducing the USDA county-level yield statistics from 1980 to 2013.
Figure S6 Maximum mean weekly vapor pressure deficit (VPD) for the Indiana farm from 1980-2013 simulated by the default APSIM method and an updated method that uses actual vapor pressure as a meteorological input (see Data S3.2 for detailed implementation).
Figure S7 Simulated daily mean canopy temperature by the STICS empirical relation algorithm vs. daily mean air temperature.
Data S1 Model review.
Data S2 APSIM description.
Data S3 Documentation for algorithms.
Data S4 Temperature response of GPP/APAR.
Data S5 Descriptions of EDD and KDD calculation.
Data S6 C4 Photosynthesis Model.